

The internal information transformation of IBM



J. A. Vayghan
S. M. Garfinkle
C. Walenta
D. C. Healy
Z. Valentin

The ability to utilize data as an enterprise asset is central to every enterprise transformation initiative. This ability is critical for reusing data consistently throughout the enterprise and deriving actionable knowledge from it. Accurate and high-quality data must consistently propagate meaning and value throughout the enterprise and comply with the policies and processes of the enterprise. For a variety of reasons, large enterprises manage data at a local level (e.g., for each department and location), resulting in information “silos” where data is redundantly stored, managed, and processed, each with its own policies and processes, leading to inconsistency. IBM has begun a transformation process to establish a program for the management of its critical data, beginning with the creation of an enterprise data strategy that is aligned with IBM business strategy. In this paper, we describe the progress, to date, of the IBM transformation process. We focus on the activities of the IBM Enterprise Business Information Center of Excellence (EBI CoE), which is responsible for establishing, implementing, and deploying the enterprise data architecture program. The EBI CoE uses leading-edge information management technology and services from IBM and its partners to address enterprise data challenges. We present lessons learned and best practices derived from this ongoing internal transformation process that can be useful for enterprises facing similar data challenges as they transform their operations and business models.

INTRODUCTION

Global technological, political, and social changes are removing the divisions that separate companies, governments, and people. Large enterprises, including governments and companies, face pressures in a number of areas, including globalization and disaggregation of the enterprise, increased external and internal regulations and compliance requirements, mounting global competition and resulting cost pressures, and massive data growth. These

trends often result in data inconsistency, duplication, and concerns for security and privacy.

In response to these increasing pressures, enterprises need to replace the traditional “information

©Copyright 2007 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the Journal reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to republish any other portion of the paper must be obtained from the Editor. 0018-8670/07/\$5.00 © 2007 IBM

“silo” approach to data management with an integrated management system and architecture for data, processes, and policy. Data must be made available for reuse throughout the enterprise. Making the data available to be used outside legacy and monolithic applications is the first step toward creating “flat” enterprises in which data is freely and securely accessible to any user with access permission.

The internal transformation of IBM into an on demand business includes an integrated enterprise data strategy. The major components of this strategy, which enable the reuse of critical data, are: creating enterprise data services for all information domains; establishing a proactive data-quality program; implementing a centralized data governance program through an information stewardship program; implementing common processes for the creation and maintenance of enterprise data domains; creating and communicating a program that fosters an enterprise culture supportive of the enterprise information architecture; and developing the skills necessary to implement this strategy.

From a technical perspective, the integrated enterprise data strategy brings service-oriented architecture (SOA), information architecture, and event-driven architecture principles together to provide information on demand by creating reliable and secure trusted data services. At the policy level, the strategy establishes a data governance program to implement business rules and ensure regulatory compliance. At the process level, it defines repeatable and measurable processes for the creation and maintenance of all core enterprise data entities.

The office of the IBM chief information officer (CIO) has established integrated enterprise-wide programs for data architecture, application architecture, and process architecture. To implement the transformation in the data area, the EBI CoE was established. The following sections describe major challenges in the IBM transformation and details of how those challenges have been addressed by the EBI CoE.

THE INFORMATION SILO PROBLEM

Currently, many enterprises do not have a comprehensive view of their policies, are not aware if processes are adhered to, and have inconsistent information that is widely dispersed. Data entities are managed in fragments and are stored in

monolithic applications or information silos. The information silos are created and perpetuated by a “silo mentality,” and their existence can easily result in an enterprise data landscape like the one depicted (in part) in *Figure 1*. In the figure, each rectangle represents a distinct system or application. Silos can be eliminated by a solution that addresses organizational, procedural, cultural, and skill-related barriers as well as those related to information technology (IT).

Many organizations do not realize the extent of the problems created by information silos until they face the complex challenges that silos cause. For example, during acquisitions and mergers, enterprise-critical data such as customer, revenue, and sales opportunity information needs to be integrated to establish a profitable outcome. If an enterprise is selling part of its business, it can be very challenging to determine what part of its enterprise-critical data needs to be separated and transferred to the buyer to comply with all government and business policies. Information silos cause difficulties when there is a need for accurate and actionable reports and information to be available to all levels of management responsible for operations and planning, or when fact-based analytics are used to forecast market demands, market opportunities, revenue, and resources.

Information silos also make it very difficult to optimize supply chains when operations and manufacturing responsibilities are distributed among many locations and partners. Using the Internet as the main platform for interaction with customers, partners, and suppliers requires integration of information, functions, and processes that are often trapped in information and organizational silos.

To address these and other daily business challenges, we focused on solving a number of data-specific challenges:

1. *Trusted data sources must be established for enterprise-critical data*—This becomes a major challenge when ownership of data is fragmented in many areas, when there are multiple sources for the same data (each with its own unique definition), when existing data is undocumented and tightly coupled inside monolithic applications, and when subject matter experts in these areas are in short supply.

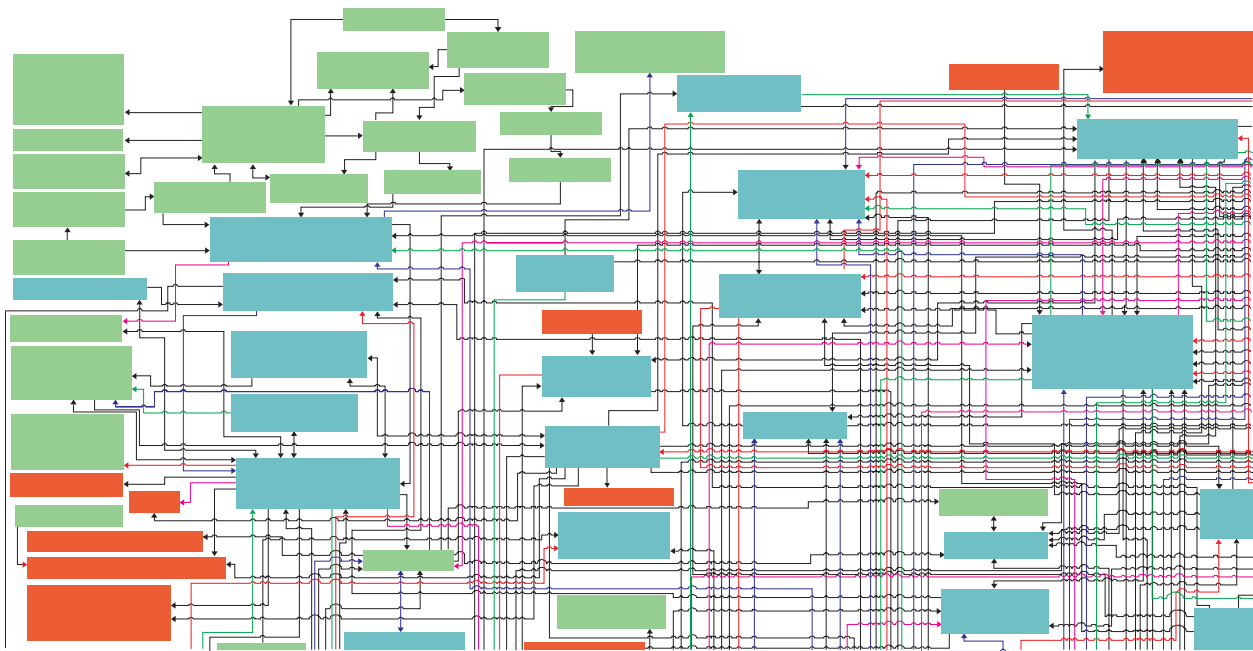


Figure 1
Typical enterprise data landscape

2. *Information must be integrated*—This includes the integration of data from multiple domains to support critical business functions in addition to business intelligence initiatives. The main contributing factors to this challenge are platform dependencies, organizational issues, tight coupling with business processes, and the lack of business data standards and an enterprise data model to define an explicit relationship among entities from multiple domains.
3. *Data quality must be improved*—This is a major challenge in any enterprise with a large amount of data in which data fields are not standardized, multiple ways of capturing the same data are used, and data is copied in many places. The increasing dependency of every enterprise on data and the increasing number of corporate and government regulations requires an integrated and enterprise-wide approach to information security and data privacy.

An industry-based specification needs to be established for information exchange among applications inside the enterprise and with external partners and suppliers. Lack of a single, common, industry-based messaging specification makes enterprise application integration a very difficult and expensive task.

A culture and an organizational structure must also be established to support broad access to data in information silos for reuse throughout the enterprise.

Traditionally, organizations have used data modelers and database programmers to solve data problems. Solving the enterprise data challenges discussed here requires new skills to address architectural, integration, security, enterprise, and data-quality challenges. Technical data strategy and business strategy must be synchronized to define a road map for managing data at the enterprise level instead of the local level. A data governance program must be established to ensure that policies and regulations are dealt with consistently throughout the enterprise.

When dealing with these enterprise data challenges, it is important to identify the type of data being managed. Most enterprises like IBM deal with three types of data: master data, transactional data, and historical data. *Master data* includes the core data entities of the enterprise (i.e., customer, offering, and employee data). Unlike many enterprises that do not deal with master data explicitly, the EBI CoE has been developing and deploying master data

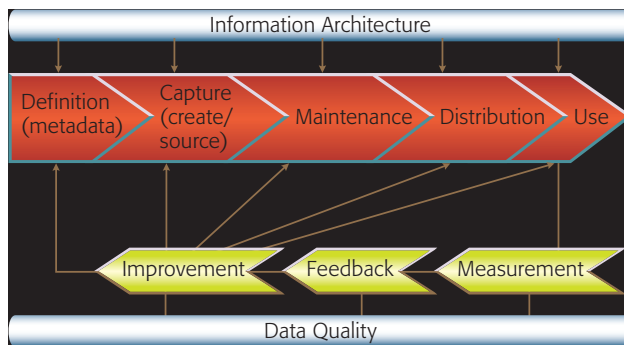


Figure 2
Typical master data life cycle

solutions for customer data and offering data within IBM. It is important to realize that each of these three data types has a unique role in the enterprise and requires its own architecture, process, and governance principles.

Each data type goes through a unique life cycle. **Figure 2** shows a typical life cycle for master data. Each phase of the life cycle has its own challenges and duration. The following section discusses a typical solution landscape and some of the major steps from the ongoing internal efforts at IBM to establish an enterprise data architecture program.

Defining a solution

Our efforts to eliminate information silos started with defining an enterprise data strategy that aligned our technical strategy with the enterprise business strategy. The main objectives of the strategy were to establish an enterprise data architecture program that would encourage the creation and use of enterprise data instead of local data.

The strategy was based on a number of principles, some of which we describe here. All data had to be of high quality in terms of its definition, business rules, and content. Worldwide business data standards had to be established and managed for data that was shared among multiple business functions, Web sites, or applications. Adherence to these standards was necessary for all architectures and associated development efforts. All worldwide business data standards required an identified business data steward who was responsible for data quality, business requirement suitability, definition, support, and use throughout IBM.

All data required a source which was authoritative for the intended information consumer. There had to be a common process for creating and maintaining shared data. Data could not be maintained redundantly unless there was a valid business justification for doing so. The data needed to be understandable, visible, and easily accessible throughout the enterprise for all data consumers.

The migration of data from one application solution to another needed to follow a disciplined approach to ensure that high quality was maintained during the transformation process. All data needed to conform to identified (internal and external) security and data privacy standards, policies, and regulations.

The enterprise data strategy and enterprise data program cover structured and unstructured data. While IBM has a comprehensive family of products and services for handling and processing unstructured data,¹ this paper does not describe our work in the area of unstructured data.

The implementation of the enterprise data strategy and elimination of information silos could not be achieved simply by defining and implementing a technical architecture. Many other elements contributed to a viable, practical, and successful solution. The IBM internal enterprise architecture program has six elements that are closely integrated:

1. *Maturity model*—Measurement of progress in the information transformation process.
2. *Enterprise data architecture program*—Documentation of the current landscape, definition of an SOA-enabled target architecture, and a road map indicating how to migrate to that architecture from the current environment.
3. *Process*—Common processes for the creation and maintenance of enterprise core data entities.
4. *Data governance*—An integrative management system for structuring, describing, and governing data as an enterprise asset, including the implementation and monitoring of data-related policies, standards and guidelines.
5. *Culture*—An organizational culture to support the elimination of impediments to open data for reuse throughout the enterprise.
6. *Skills*—A program to train employees and develop the critical skills that are needed in the data

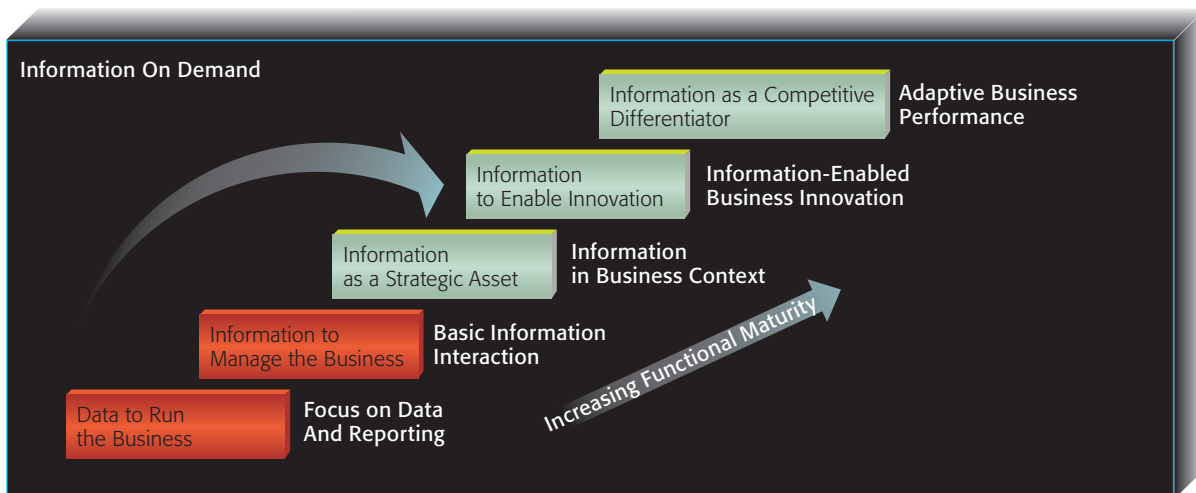


Figure 3
Business information maturity model

area to implement and maintain the enterprise data architecture program.

The remaining sections of this paper include a brief description of how the EBI CoE has approached some of these solution elements.

MEASURING PROGRESS: THE MATURITY MODEL

Measuring the maturity of business information at the enterprise and data-domain levels is critical for business transformation. Key elements in measuring the maturity of the data architecture program of an organization include methods to establish a baseline, a common and enduring model of maturity, a comprehensive means of self-assessment, and a common tool to establish a consistent set of criteria allowing each data domain to score itself. These elements enable the enterprise to conduct an internal dialog about its strengths, gaps, and actions and enhance its ability to focus on actions rather than penalties.

With this goal in mind, IBM has developed the information on demand business information maturity model depicted in *Figure 3*. It specifies five discrete levels of information integration, architectural blueprint adherence, and business intelligence capability. The model establishes a common way to conduct a dialog with customers and the owners of data inside IBM. These five levels of data management are used to assist transformation activities that improve adherence to architectural standards and

use of strategic tools, improve information integration and quality, and allow enterprise teams to create solutions that improve the competitive responsiveness of IBM to existing and emerging opportunities.

A tool has been developed that enables self-assessment of practices and the observation of data in key data elements. This is done in a way that can be compared with all other data domains inside the enterprise. Using the tool and the associated process, the “as is” environment is defined as the baseline. In addition to the baseline, the tool allows each team to set targets for improvement in the high-priority gap areas, as shown in *Figure 4*. The scoring and gap analysis identifies high-priority issues that can result in the highest level of return and benefit to the enterprise.

ENTERPRISE DATA ARCHITECTURE

This section presents a description of the technical architecture for the IBM internal enterprise data architecture program. The internal architecture blueprint is documented using standard notation and tools and is stored in an enterprise repository that is accessible over the IBM intranet.

Components of the architecture

Based on the principle that the accurate statement of a problem is a critical part of its solution, the first step in the creation of a technical enterprise architecture is to document the current landscape

<i>Business View</i>		Now	YE06	YE07	Rationale for Scoring
Total Business View		0.0	0.0	0.0	
<i>Data Requirements and Data Model</i>					
1. The logical data model is understandable and approved by senior management as representing what the enterprise must know to perform its processes and accomplish its mission.					User Guide: Key components of scoring: the degree to which the logical data model is defined, documented, and understood within the enterprise; how well the information model maps to what must be known within the enterprise for it to operate effectively and achieve its mission-critical objectives; to what degree senior management is aware of and formally approves of the information model. Examples: • Score 1 – The data model is not well defined and not documented. • Score 2 – The data model is partially defined and is somewhat known and used within the enterprise. Significant gaps exist in the model that prevent it from effectively supporting the information needs of the enterprise in order to accomplish its mission. Senior management has not approved the model. • Score 3 – The data model is fairly well defined, but improvements are needed. Senior management may or may not have approved the model. • Score 4 – The data model is defined, documented, and approved by senior management as meeting the enterprise's needs in order to accomplish its mission. Some improvements are needed to bring the enterprise from the data and information stage to exploitation of knowledge as a competitive advantage. • Score 5 – The model is complete and well defined, and undergoes validation and continuous improvement as part of enterprise planning cycles. Senior management has approved the model as a key element of the enterprise's mission achievement. The model is also used as a key element of the enterprise's knowledge management strategy.
2. Data elements self-analyze for patterns of use and self-adjust based on use.					
3. Common data elements and information models are reusable across all business processes and avoid redundancy where possible.					
4. Data and information models enable capture, maintenance, and dissemination of data among knowledge workers, even as business processes or applications change.					
5. Data and model are extensible and upward-compatible to any new requirements (i.e., are not impacted by migrations or transformations).					
6. Data content and model map to strategic business processes of enterprise or community of interest (covers data exchanges beyond enterprise boundary).					
7. Data warehouse and data-mart physical designs accurately implement information requirements as defined in strategic data/information model.					

Figure 4
Information management tool output

and profile all information assets. Analysis of the current landscape uncovers problematic areas while establishing a baseline that must be used to measure progress in the data transformation process.

The next step is to establish a target architecture to support the enterprise mission and align it with the business strategy, if one exists. The enterprise architecture team usually faces many alternatives when defining the target architecture, and should favor alternatives that reuse existing enterprise components. The team should also choose alternatives that make migration from the current environment to the target environment a practical and affordable task, while establishing standardized components that are usable throughout the enterprise. **Figure 5** depicts the major components of the IBM target data architecture blueprint. A *data-mart*, as shown in the figure, is a specialized subset view of domain-specific or focused analytical information.

A practical, affordable, and feasible road map is needed to define the transition to the target

architecture. It is important to consider enterprise capabilities and constraints when defining the road map. To ensure that the target architecture and road map will not be stored in the corporate blueprint repository without any resulting action, these components need to be affordable, publicized, and incorporated into the culture at all levels of the enterprise.

Additionally, an architecture board needs to be established to review all major enterprise initiatives, ensure that their implementation is consistent with the road map, and verify that they reuse enterprise components, create reusable components, and adhere to the target architecture. In some cases, the architecture board may request that a certain project implement one of the common components defined in the target architecture of the enterprise. The architecture board will also demand that project teams use standards and enterprise components in their development. The architecture board review process must make all necessary provisions to make sure that the local and project-specific investments gradually move the enterprise toward the target

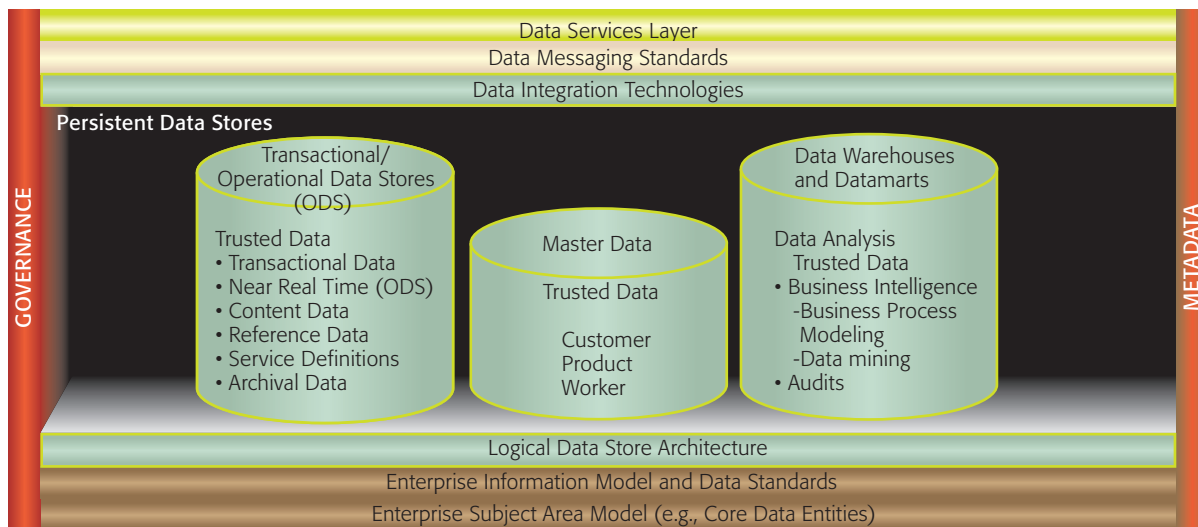


Figure 5
Components of target data architecture

architecture. The architecture board makes recommendations to the executive team, which makes enterprise investment decisions.

Since the world is continuously changing, it is imperative to review the target architecture and road map on a regular basis to ensure that they continue to define the correct direction for the enterprise.

Architecture process

In this subsection, we present a noncomprehensive list of some of the major steps of a process for defining the target enterprise data architecture.

Subject areas that are important to the enterprise, such as customer, offering, employee, ledger, and sales-order information, must be identified. An enterprise subject-area model with a hierarchical structure must be used to organize those subject areas. The enterprise subject-area model is used to identify subject areas that need to be governed by a data steward. Subject areas are also known as the *vocabulary* of the enterprise.

A logical entity relationship model must be created to clarify the relationships between subject areas as they support enterprise-critical business processes. Metadata must be defined to describe all aspects of the subject areas (e.g., data structure, processes, and rules). An enterprise metadata repository must be defined as part of the target architecture.

An industry standard must be adopted to specify a standard definition and implementation of data exchange between applications within the enterprise and with external partners. The EBI CoE selected the Open Application Group Integration specification² as a base for the definition of a canonical business language for information exchange within the enterprise and with external partners. Core enterprise data entities must be identified and incorporated into common components as part of the enterprise master-data initiatives.³ The target architecture needs to include these common components with necessary data services. Examples of these entities are customer, offering (e.g., product or services), and employee data entities.

A data integration layer must be included to combine the data that is distributed in many local repositories. The purpose of this layer is to create a consistent view of data that is being managed as fragments within the enterprise.⁴ A data services layer must be included to decouple data consumers from the manner in which data is structured and stored.⁵ The IBM Service-Oriented Modeling and Architecture (SOMA) is used to identify the data services.⁶ Common components must be defined to support information security and privacy requirements.⁷

As discussed in the previous sections, an enterprise has three types of data, each having a distinct life

cycle. The target architecture and its definition process need to address each data type and its corresponding life cycle appropriately.

Architectural style

The enterprise target data architecture has a hybrid architectural style, bringing three styles together to establish the foundation for an agile, loosely coupled, on demand enterprise: enterprise information architecture, SOA, and event-driven architecture.

The enterprise information architecture defines trusted sources of information, metadata, business data standards, data quality, information integration, and information aggregation techniques for the enterprise.

The establishment of trusted data services depends on knowing where to find trusted sources of information. Metadata consolidates explicit information about the location, structure, context, and usage of data. The metadata is essential to the creation of metadata services that are needed to establish discoverable trusted data services for the enterprise. Data standards are required to establish a forum for resolving differences in the meaning of data and the implementation of a proactive data-quality program. Data integration techniques are used to establish a common view of data. Business intelligence depends on the existence of a mechanism for data aggregation and the transformation of aggregated data into meaningful and actionable business information. An integrated governance program for data and the SOA is also an important part of the enterprise information architecture.

The SOA provides an effective foundation to integrate data with applications and processes by creating loosely coupled components. The SOA enables the delivery of information as a service by providing one-to-one communication between system components, consumer-based triggers, synchronous operations, and standard specifications for information exchange.

The event-driven architecture provides a foundation for the establishment of asynchronous operations between system components. Event-based triggers, such as those based on publication or subscription, are used to establish many-to-many communications between system components.

The application of SOA principles to enterprise data introduces new technical and business challenges. For example, in a monolithic approach to application development, all security and data privacy requirements are handled at the application level; when data is made available in an SOA platform and through data services, a new design and architecture for security and privacy at the data service level are required. An adequate level of SOA governance is the key to the success of any SOA project, even when SOA is applied to the enterprise data architecture. However, to prevent the creation of multiple governance programs, it is imperative to integrate SOA and enterprise data governance programs with other governance processes within the enterprise.

Trusted data services versus trusted data stores

Without establishing and maintaining trusted data, the enterprise cannot take advantage of the value of information. Establishment and utilization of trusted data is a cornerstone of the IBM internal transformation efforts. The transformation strategy depends on information that originates from available trusted sources. The trusted data is accessed as a service through trusted data services, rather than requiring data consumers to access trusted data stores directly (i.e., by using a point-to-point connection) and establish dependencies on the database structure.

In a typical environment, with direct point-to-point connections between trusted data stores and the data consumer, the responsibility for data ownership and data utilization decisions is shared among multiple groups. The database owner is typically responsible for maintaining the data and data consumers are responsible for creating their own specific interfaces to access the data from the database. When data is used by many groups, many interfaces to the same data are created, each with a unique view and interpretation of the data. This is one of the reasons why data may be interpreted in many ways within an enterprise. In contrast, in an SOA-enabled enterprise data environment, where the data is provided as a data service, the data owner is responsible for providing the trusted data services. In this case, the same data service is used by all data consumers. Thus, the chance of the same data being used and interpreted differently is minimized. This change in responsibility may require organizational adjustments in the enterprise.

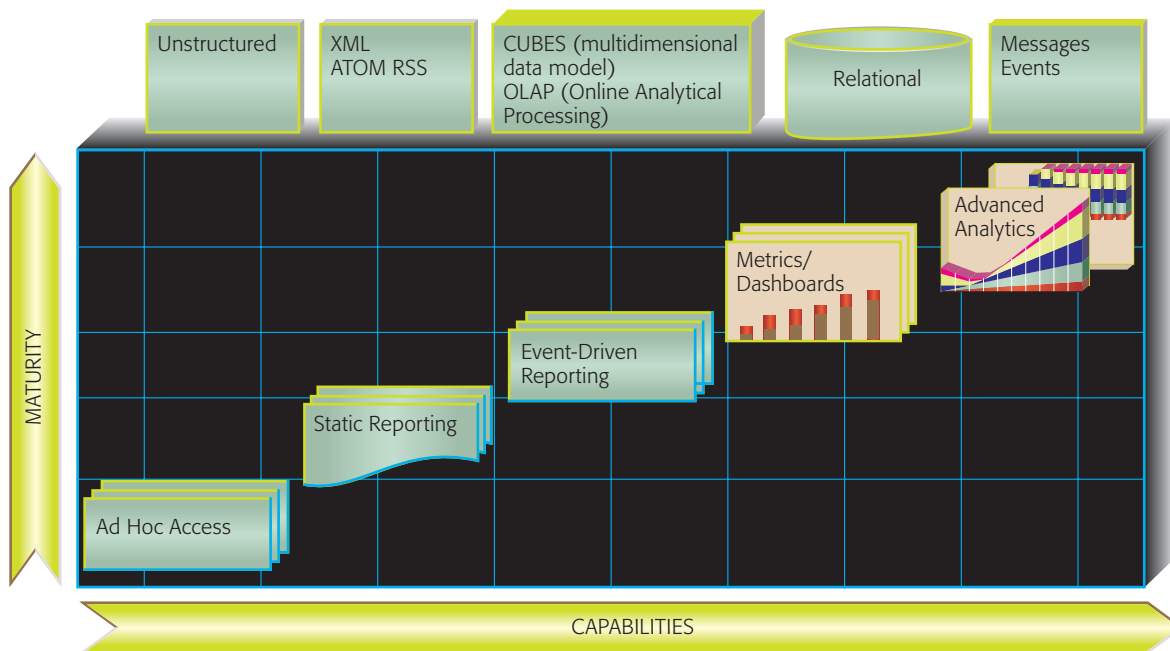


Figure 6
Value of business intelligence

Information security

Service interruptions, unauthorized access, stealing and altering information, identity theft, virus proliferation, theft of backup tapes, and fraud are just a few of the security risks in an enterprise.⁸ Security and data privacy are extremely important aspects of business transformation and require special attention. The growth of the Internet, replication of data, and requirements for ease of access all contribute to the complexities of securing information assets. The enterprise must ensure that all internal and external mandates for standard and regulatory compliance are incorporated into the information delivery process and the enterprise data architecture program. When data services access data directly and thus bypass authorization functions that were implemented at the application level, unauthorized use of data must be prevented. Other elements for consideration in securing the information enterprise include the ability to support firewalls, protect intranets, and traverse the Internet.

The target enterprise data architecture provides an integrated approach to information security by using authentication and authorization functions. Authentication is the process of verifying a user's true identity; entitlement or authorization involves

identifying the roles and policies associated with the required resource access control.

IBM has deployed a common authentication system that is used to establish a common and consistent authentication for all existing enterprise data and applications. To create a common, role-based authorization and entitlement processing system for the emerging SOA-enabled data architecture, the EBI CoE uses IBM WebSphere® DataPower SOA Appliances and Tivoli Access Manager (TAM).

This solution strategically utilizes the capabilities of each organization to collect authorization rules, verify entitled users, and provide methods for enforcing security policies in a common way.

Business intelligence

Decision-making assistance and business intelligence can have a significant impact on the growth of a business.⁹ The business value of information depends on how it is used to support business activities. Stored data and information cannot be considered as an enterprise asset unless they are used to influence and guide business activities. Data provides factual information, but it also drives business direction and investment. *Figure 6* shows

the increasing business value of information management (IM) capabilities as defined by the EBI CoE.

The EBI CoE business intelligence team's mission is to enable the IBM on demand transformation to obtain the right information when and where it is needed for business insight and performance management. To fulfill that mission, a leading-edge business intelligence system called EDGE (Enabling Decisions for Global Execution) on Demand has been developed and deployed. The EDGE on Demand system is designed for use by executives, managers, and professionals in support of the internal management system and day-to-day activities and processes of IBM.

The EDGE on Demand system provides standard reports that support the execution of business management systems and standard processes: reliable, accurate and on-schedule report production; controlled access to all reports and data based on business need; report development and maintenance services to enhance the report portfolio in concert with rapidly changing and complex business needs; "point in time" data for all reporting from trusted data sources; and consistent sales management reports for all brands, markets, channels and geographies, for all management levels. These features deliver benefits such as: 1) increased individual and team productivity by allowing users to focus time and energy on analysis and action, rather than report preparation or reconciliation; 2) improved global business and cross-business operations; and 3) reduced overall expenses associated with reporting. The EDGE on Demand system received the top honor in the 2005 Best Practices in Data Warehousing competition of the Data Warehousing Institute.¹⁰

Additionally, IBM has been very active in the application of advanced analytics to manage its sales and marketing operations. The advanced analytics capabilities support and drive several major business transformation projects. Quality business information, when delivered securely, accurately, and on time, enables the enterprise to compete and grow by making actionable insight, the basis of business intelligence, possible.¹¹ Gathering, managing, analyzing, and distributing data throughout and outside of the enterprise can result in making better-informed decisions.

The quick delivery of analytics and information to new and existing customers is also imperative. Without the ability to deliver needed information on demand, informed business decisions cannot be made when needed. Placing information on the Web engenders new expectations, including the need for real-time alerts when the data has changed. Information feeds such as the Atom Publishing Protocol (APP) and Real Simple Syndication (RSS) provide convenient methods for end users to ensure that they receive only the most current content and only the content in which they are interested. The enterprise must determine which information assets should be made available in the form of live feeds, static information, or through transactional Web services.

PROCESSES

The transformation of processes is a critical part of the internal transformation of IBM. Establishing enterprise-wide processes for the creation and maintenance of master data is an important element of the IBM enterprise data strategy. The EBI CoE has the responsibility for defining and deploying those for the customer and product domains. This section includes a brief overview of a common worldwide process that the EBI CoE has defined for enterprise management of customer data. This includes the definition of a common worldwide process for creating, maintaining, and managing customer data; the establishment of requirements for the adoption and deployment of the common process through user activities and technology; the establishment of a worldwide process management system that includes process ownership, compliance, measures, and controls; a means to ensure compliance with data standards; and a means to methodically establish linkages to critical IBM transformation efforts to guarantee continuing process efficiency.

Adopting a common process ensures that regardless of the organization or application, all users will go through a common set of steps in managing customer data. This begins with enabling a strategic single view of the customer and simplifying data exchange, thereby reducing overall cost, adding flexibility with staffing and support operations, and creating opportunities to automate some of the process.

The enterprise information process approach transforms customer information incrementally. This

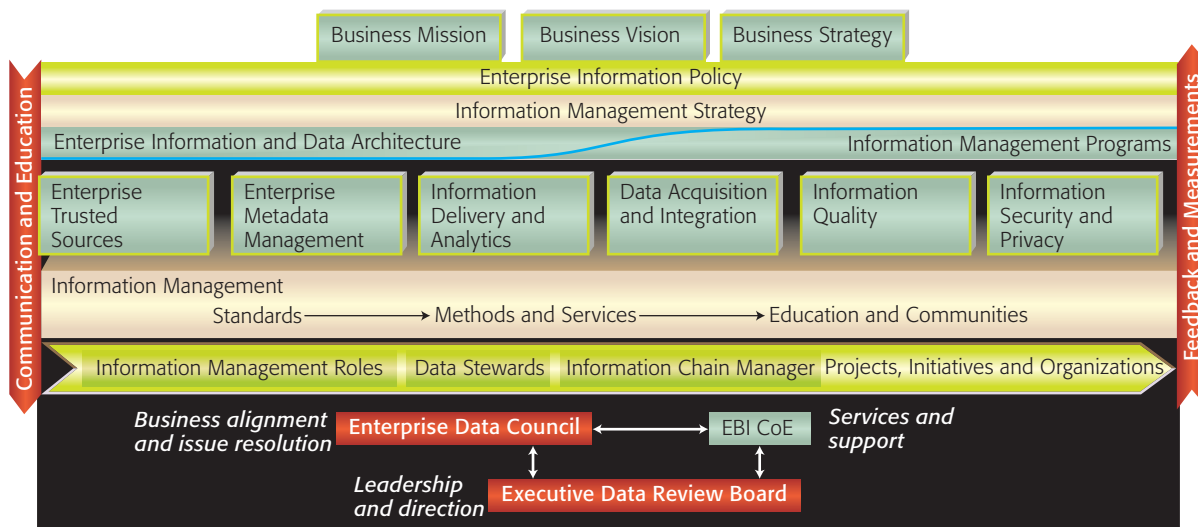


Figure 7
Data governance at IBM

approach instills a practical discipline that does not attempt to make abrupt global changes, but rather focuses on moving strategic capabilities through levels of maturity. As priorities or business climates change, process priorities can adapt to meet new business needs.

Business standards for customer data are also defined as part of this process. The process standards for customer-information data management focus on the following criteria: an overall reduction in turnaround time through streamlining, automation, and simplification of the creation and updating of customer records; reliance on data standards for processes, which fosters improvements in data quality and consistency and leads to increased efficiency in moving data to other systems; and utilizing technology to take advantage of existing IBM and industry “best of breed” solutions, including IBM Customer Data Integration solutions.¹² Using these proven solutions reduces overall development and deployment time.

The processes implemented within the enterprise ensure that today’s successful business relationships—whether with customers, partners, or suppliers—focus on integrating all of the information an organization has gathered about the customer and analyzing it to gain valuable insights. The information is then put into the hands of the people who make the day-to-day decisions that most affect the customers of that organization.

DATA GOVERNANCE

Within the information environment, governance applies to many aspects of data management. The enterprise data architecture program and its associated processes include data governance (DG), an integrative discipline for structuring, describing, and governing information as an enterprise asset without regard for organizational and technological boundaries. It involves implementing, monitoring, and sustaining information policies, standards, and guidelines in order to improve operational efficiency, promote transparency, and enable data quality control, regulatory compliance, business information maturity, and business insight.

At IBM, DG is used to provide information policies and guidance, enabling the deployment and application of these policies through a set of IM practices. A feedback mechanism is provided through compliance processes, selected DG or IM metrics, and data governance forums, as shown in *Figure 7*.

The DG structure must be integrated with enterprise governance structures in other areas. Some of these areas include regulatory compliance, governance of IT and transformation investments, governance of IT infrastructure and enterprise architecture including SOA, governance of process deployments, governance of business controls, and risk management. The DG structure is also linked to the overall business vision, mission, and strategy. The business and general environment needs to be taken into

consideration when deciding how, and in what sequence, to implement the DG components, processes, and principles.

While the model may suggest a “top down” approach to DG, in reality, parts of the DG model may be implemented at different times, thus making it a hybrid approach with some “top down” activities and some “bottom up” activities. Often an executive leader or sponsor emerges from a selected business group, developing and validating some of the DG concepts before promoting an enterprise-wide view and positioning DG overall. These executive sponsors are critical to obtaining and sustaining enterprise-wide support before enterprise-level governance is implemented.

Components of DG

The organization must have a formal set of directives and policies for the management of information within the business. An information policy is best implemented in the same way other corporate policies are handled. This facilitates employees’ recognition of the information policy and enables the incorporation of information directives into existing compliance and audit processes and structures.

The information policy is based on and supported by a related enterprise IM strategy, which holistically describes the way the organization manages information. This strategy addresses questions such as: What processes are used to manage information through its life cycle? How does data stewardship influence IM? What is the business impact and value of managing information and the cost to the business if information is not managed appropriately? How are information priorities determined and funded? How are information issues and dependencies managed? What is the difference between the enterprise view of IM and the business-unit or functional view? How can it be determined if the enterprise information is effective for the business? What information metrics should be managed as key performance indicators (KPIs)? How does the business ensure ongoing and consistent execution of information processes?

The next level of data governance focuses on IM practices or programs and is referred to at IBM as the information management framework. In each of the IM programs, a set of standards and guidelines is

provided that aligns with the overall IM policies. In addition, the IM program should provide the definition or reference to the correct methods, practices, technologies, and tools to be used, ensuring that the policies and standards can be implemented.

There are many different ways to organize IM practices. In the absence of industry standards, the EBI CoE has categorized the IM programs in the following way: The category *enterprise trusted services* includes master data management and data warehousing. *Enterprise metadata management* includes, among other things, business information modeling, business data standards, and messaging standards. *Information delivery and analytics* determines the usage of data through reporting or business analytics. *Data acquisition and integration* includes data provision, transformation, extraction, loading, and migration. Finally, the categories *information quality* and *information security and privacy* ensure data accuracy and protection.

It is very important to define, communicate, and deploy the appropriate standards, methods, and tools for each IM program. A prerequisite to the deployment of IM programs is the definition and creation of IM roles, such as information architects, information-quality analysts, data modelers, and data stewards. The enterprise must provide adequate educational programs for the ongoing enhancement of employees’ skills and knowledge. Ultimately, this can result in improved career paths and professional certifications for IM personnel, and retention of critical skills for the enterprise.

As business and IT teams apply these methods and standards in their transformation projects and in life cycle management work, the enterprise DG team needs to determine how it interacts with these teams to ensure compliance and provide support as required. The EBI CoE has established an IM system offering assessments and advisement to teams in their implementation efforts, as well as providing a feedback loop to the IM program office to foster success and further improvements of the methods and standards. Parts of the IM system are fully integrated with other governance processes in IBM that ensure architectural compliance and approval of new IT investment decisions. This integrative approach ensures that data, applications, and

infrastructure are effectively treated as equal elements in the internal business transformation.

Data stewardship is a key element in enabling the DG process, and is defined as the formal accountability for the management of data as an effective enterprise asset. There are different ways of implementing data stewardship, which usually include establishing accountabilities at various levels of the organization, such as the executive level, managerial level, and knowledge worker (or subject matter expert) level. Data stewardship can be organized by the type of information (e.g., subject areas) involved, the end-to-end business processes, or both. Stewardship often crosses organizational boundaries. Data stewardship responsibilities can vary, depending on the organizational approach to implementing IM processes; however, they usually include the role of a data steward in conjunction with other IM roles, such as information architect, data modeler, or data-quality analyst. An organization may choose to define different stewardship roles to differentiate between levels of scope, authority, and responsibility (e.g., tactical steward, operational steward, executive steward, steward delegate, or data owner). A central program office can be used to implement the data stewardship program consistently throughout an organization or enterprise.

IBM implemented an Enterprise Data Council (EDC) for the data stewardship teams and an Executive Data Review Board (EDRB) for transformation executives to manage IM initiatives, priorities, governance, and issues. The role of the governance bodies is to approve IBM information policies, set the enterprise-level information priorities, monitor the overall effectiveness of the business information, generate business value through IM programs, resolve enterprise-level data issues, and ensure IM accountabilities throughout the business. The EDC represents the data stewardship teams and working groups, while the EDRB is the body responsible for decision-making and final-issue resolution. Together, they manage the overall enterprise business information maturity process.

Implementation of DG

There are many ways to implement a DG system. While it may seem logical to start with the definition of the enterprise information policies without underlying methods and practices for their compli-

ance and implementation, policies defined in this way are often not effective, and can be counterproductive. In a worst-case scenario, the entire enterprise can rapidly become noncompliant.

Implementation of the DG agenda may be encouraged by other governance needs. For example, establishing an SOA or complying with the Sarbanes-Oxley Act (SOX) may be the impetus for managing information as an asset.

Expanding DG from single business organizations to a truly enterprise-wide system requires strong executive leadership and persistence in overcoming challenges along the way. Some of the typical challenges in this context include overcoming parochial views and priorities from business functions that prevent enterprise-level adoption. An enterprise-level funding and accountability model should be established, allowing functional business leaders to play their enterprise roles. A corporate-wide DG or IM office is required to lead IM strategy, program development, and deployment. There must be sufficient investment in skills development, education, and career paths for IM professionals. IM metrics and KPIs must be included in the enterprise measurement system and business objectives. IM issues and data-quality problems must be brought to prominence to raise awareness and make the case for improved IM and DG processes.

At IBM, we have addressed these challenges in various ways. Critical to success in these efforts was an executive sponsor who advocated the need for and importance of good IM practices at the senior-management level. One effective strategy has been demonstrating successes by resolving known data issues in selected business areas. Success stories act as proofs of concept and provide experience for further refinement of the IM practices. One team's success inspires other teams and organizations to begin adopting the same best practices. This was further encouraged through introducing IM reviews of projects into existing IT governance processes, requiring project teams to obtain approval from the IM teams.

The realization of the importance of robust IM practices occurred at all levels of the organization, a necessary element for persistent culture change. Cross-functional forums such as the Data Governance Council or communities of practice had an

influence on teams and peers, further breaking down functional silo thinking and encouraging the acceptance of enterprise-wide IM. This has been a multiyear process, and most organizations within IBM are still undergoing the transformation initiated by adopting IM best practices.

The sponsors of enterprise DG need to find ways to determine and articulate the overall business case for IM. Without IM, the business incurs added costs due to poor data quality, redundancy in data systems, and inefficient workflows resulting from information unavailability. In addition, there is a risk of lost marketplace opportunities and of customer satisfaction issues caused by a lack of business insights and analytics. The ultimate goals of these IM disciplines and practices are to enable and optimize more effective business processes.

CULTURE

As discussed earlier, the successful transformation to an on demand enterprise in which trusted data is provided as a service requires the removal of information silos. This can only be achieved by the elimination of the silo mentality and narrow vision within the enterprise. When this happens, the culture of the organization is changed. Information silos collapse, and data is made available for reuse throughout the enterprise.

Traditionally, this culture change has been gradual, beginning slowly but accelerating if some early successes with tangible business value can be demonstrated. An example of these successes would be the establishment of a set of data services for critical data and their use with emerging Web 2.0 technologies to solve critical business problems that otherwise would have required development of a multiyear, multi-million-dollar monolithic application.

The enterprise approach to data management needs to have consistent and strong support from senior executive teams. Continuous execution of DG and establishment of the IM profession will force cultural changes and accelerate them as well.

CONCLUSION

Transformation to an on demand business requires the elimination of information silos that trap data in local applications and processes. The removal of

silos can make actionable and high-quality information available on demand. This transformation is a process that cannot be completed by simply deploying a new technology, but requires removal of silo vision and silo thinking. Other components of the transformation involve changes in processes, organization, governance, and culture.

The transformation solution is derived from an integrated enterprise data strategy that aligns technical strategy with the enterprise business strategy to enable business integrity at the enterprise level. The ultimate goal of the enterprise data architecture program is to liberate data that has been confined in legacy and monolithic applications and make it available for reuse while complying with all relevant policies and rules.

The internal enterprise data architecture of IBM brings together concepts and principles from information architecture, SOA, and event-driven programming to establish trusted data services that make information available on demand. Liberating data confined in information silos and making it available through trusted data services creates new challenges in data security and data ownership. The IBM EBI CoE has implemented a DG program and repeatable processes for the creation and maintenance of master data in order to deal with these challenges.

The enterprise information transformation and establishment of an enterprise data architecture program for IBM is an ongoing process that enjoys consistent support from senior executives and from technical and business teams. Lessons learned and best practices from this process are used to develop products and services that will help clients who face similar enterprise data challenges.

ACKNOWLEDGMENTS

We would like to thank the IBM Enterprise Business Information Center of Excellence employees and all of our colleagues and clients inside and outside of IBM who have contributed to and supported the internal information transformation of IBM. Special thanks go to Kathy Price, Joe Balducci, Brad Goss, Sunita Menon, and Pamela Dugan for lending us their expertise.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

CITED REFERENCES

1. *Enterprise Content Management*, IBM Corporation, <http://www.ibm.com/software/data/cm/>.
2. M. Rowell, *The Open Applications Group Integration Specification*, IBM Corporation (June 2003), <http://www.ibm.com/developerworks/xml/library/x-oagis/>.
3. *IBM Multiform Master Data Management*, IBM Corporation, <http://www.ibm.com/software/data/masterdata/>.
4. *Information Integration*, IBM Corporation, <http://www.ibm.com/software/data/integration/>.
5. *Information On Demand: Leveraging the Business Value of Information*, IBM Corporation, <http://www.ibm.com/software/data/information-on-demand/>.
6. A. Arsanjani, *Service-Oriented Modeling and Architecture: How to Identify, Specify, and Realize Services for Your SOA*, IBM Corporation, <http://www.ibm.com/developerworks/library/ws-soa-design1/>.
7. *Security Management*, IBM Corporation, <http://www.ibm.com/software/tivoli/solutions/security/>.
8. C. Steel, R. Nagappan, and R. Lai, *Core Security Patterns*, Prentice Hall PTR, Upper Saddle River, NJ (2005).
9. L. S. Sanford and D. Taylor, *Let Go To Grow*, Prentice Hall PTR, Upper Saddle River, NJ (2005).
10. The Data Warehousing Institute, <http://www.tdwi.org/>.
11. M. Biere, *Business Intelligence for the Enterprise*, IBM Press, Indianapolis, IN (2003).
12. *Customer Data Integration*, IBM Corporation, <http://www.ibm.com/software/data/masterdata/cdi.html>.

Accepted for publication May 21, 2007.

Published online November 6, 2007.

Jamshid A. Vayghan

IBM Enterprise Business Information Center of Excellence (EBI CoE), 650 3rd Ave. South, Minneapolis, MN 55402 (vayghan@us.ibm.com). Dr. Vayghan is the IBM chief enterprise data architect and manager of the EBI CoE Innovation and Emerging Technology team. He is a Senior Technical Staff Member and certified as a Senior Executive IT Architect by IBM and as a Master Certified IT Architect by the Open Application Group. He has a B.S degree in chemical engineering and M.S and Ph.D. degrees in computer science. He is also an adjunct faculty at the University of Minnesota and the University of St. Thomas. Dr. Vayghan is a recognized expert in data mining and enterprise application and data architecture. His work in the software industry has dealt with both embedded and enterprise systems, with a focus on bridging the gap between technology and business. His research, teaching, and professional engagements are in enterprise architecture and development, software engineering, database systems, data mining, enterprise SOA, and enterprise mash-up development. He is a coauthor of book chapters and papers on data mining and database systems.

Steven M. Garfinkle

IBM Enterprise Business Information Center of Excellence (EBI CoE), 2070 Route 52, Hopewell Junction, NY 12533 (garfinks@us.ibm.com). Mr. Garfinkle is a senior enterprise information architect and a member of the EBI CoE Innovation and Emerging Technology Team. He is a Senior Technical Staff Member and certified as a Senior Executive IT Architect by IBM and as a Master Certified IT Architect by the Open Application Group. He has led the technical and architectural aspects of many enterprise-critical initiatives, and has

invented solutions and processes that help to develop and leverage IBM products. Mr. Garfinkle has filed several patents on security, business intelligence, and services. He was integral to earning a Data Warehousing Institute award for business process management through his architectural leadership and innovation.

Christian Walenta

IBM Corporate Headquarters, Enterprise On Demand Group, 2070 Route 52, Hopewell Junction, NY 12533 (walenta@us.ibm.com). Mr. Walenta has held a variety of professional and management positions with IBM in Germany, the United Kingdom, and the United States. His professional experience of more than 20 years includes information management (IM), data warehousing, management of supply chain systems, business development, and project management. He is a recognized leader in IM at IBM and currently heads the IBM Enterprise Information Management office, which is responsible for providing IM strategy, programs, and data governance throughout the enterprise. Mr. Walenta developed and implemented information quality programs for the IBM supply chain and has extensive practical experience in all aspects of IM, including setting overall IM strategies, implementing information stewardship and governance programs for complex organizations, and driving IM as a continuous culture change. He serves on the board of directors of the International Association for Information and Data Quality (IAIDQ) and has been a speaker on information quality management at conferences. He is a certified IBM Executive Project Manager and holds an M.S. degree in computer science and business administration, as well as an M.B.A. degree in international and strategic management.

Donald C. Healy

IBM Corporate Headquarters, Enterprise on Demand Group, 294 Route 100, Somers, NY 10589 (donh@us.ibm.com). Mr. Healy is the program manager of the Enterprise Business Information Center of Excellence project office. This project office is a matrix team of subject matter experts with backgrounds in data architecture and modeling, information management best practices, and business intelligence solutions. The focus of this new project team is to provide consulting and perform services that will improve the maturity of IBM data management. He brings 28 years of IBM experience in large systems, software business planning, pricing, and IT strategy to this role.

Zulma Valentin

IBM Corporate Headquarters, Enterprise on Demand Group, 1133 Westchester Ave., White Plains, NY 10604 (zvalent@us.ibm.com). Ms. Valentin is a senior manager responsible for information stewardship programs for the customer and product data domains. She joined IBM in 1984, working with banking and finance customers, and in market intelligence, where she worked on database marketing. She moved to Enterprise Business Information in 1999, where her background with data quality and data management is utilized in data-quality initiatives, data governance, and business data standards. She has a B.B.A. degree in business and an M.S. degree in organizational development. ■